Volume 13 • Number 5 • 2010
VALUE IN HEALTH

# A Portuguese Value Set for the SF-6D

Lara N. Ferreira, PhD,[1,3] Pedro L. Ferreira, PhD,[2,3] Luis N. Pereira, PhD,[1] John Brazier, PhD,[4] Donna Rowen, PhD[4]

[1]School of Management, Hospitality and Tourism, University of the Algarve, Faro, Portugal; [2]Faculty of Economics, University of Coimbra, Coimbra, Portugal; [3]Centre for Health Studies & Research, University of Coimbra, Coimbra, Portugal; [4]Sheffield Health Economics Group, School of Health and Related Research, University of Sheffield, Sheffield, UK

## ABSTRACT

**Objectives:** The SF-6D is a preference-based measure of health derived from the SF-36 that can be used for cost-effectiveness analysis using cost-per-quality adjusted life-year analysis. This study seeks to estimate a system weight for the SF-6D for Portugal and to compare the results with the UK system weights.

**Methods:** A sample of 55 health states defined by the SF-6D has been valued by a representative random sample of the Portuguese population, stratified by sex and age ($n = 140$), using the Standard Gamble (SG). Several models are estimated at both the individual and aggregate levels for predicting health-state valuations. Models with main effects, with interaction effects and with the constant forced to unity are presented. Random effects (RE) models are estimated using generalized least squares (GLS) regressions. Generalized estimation equations (GEE) are used to estimate RE models with the constant forced to unity. Estimations at the individual level were performed using 630 health-state valuations. Alternative functional forms are considered to account for the skewed distribution of health-state valuations.

**Results:** The models are analyzed in terms of their coefficients, overall fit, and the ability for predicting the SG-values. The RE models estimated using GLS and through GEE produce significant coefficients, which are robust across model specification. However, there are concerns regarding some inconsistent estimates, and so parsimonious consistent models were estimated. There is evidence of under prediction in some states assigned to poor health. The results are consistent with the UK results.

**Conclusion:** The models estimated provide preference-based quality of life weights for the Portuguese population when health status data have been collected using the SF-36. Although the sample was randomly drowned findings should be treated with caution, given the small sample size, even knowing that they have been estimated at the individual level.

*Keywords:* health-related quality of life, preference-based measure, SF-6D, system weight.

## Introduction

Economic evaluation involving cost-effectiveness analysis is increasingly being used to inform resource allocation in health care. Cost effectiveness analysis using cost-per-quality adjusted life-year (QALY) analysis enables comparisons across a wide range of diseases and treatments using a common measurement. The "quality" part of the QALY is estimated using a preference-based measure of health. This preference-based measure produces a single preference-based index that can also be used to enable comparisons between socioeconomic groups and the investigation of the magnitude of health status utility differences.

The SF-36 is one of the most commonly used measures of health-related quality of life (HRQoL), yet cannot be used in economic evaluation in its existing form because it cannot generate a single preference-based index. The SF-6D is a recent preference-based, indirect utility assessment instrument created by Brazier et al. [1] and is derived from the SF-36. The SF-6D enables a utility score to be generated from the SF-36 for use in cost-per-QALY analysis. Brazier et al. [1] produced a set of preference weights for the UK general population but non-UK populations may have different preferences to non-UK populations. This study seeks to estimate a system weight for the SF-6D for Portugal and to compare the results with the UK system weights. In this article, we present the valuation survey undertaken with the SF-6D and the modeling results.

*Address correspondence to:* Lara Noronha Ferreira, Adjunct Professor, Escola Superior de Gestão, Hotelaria e Turismo, Universidade do Algarve, Campus da Penha, 8005-139 Faro, Portugal. E-mail: Lnferrei@ualg.pt
10.1111/j.1524-4733.2010.00701.x

### The SF-36

The SF-36 is a generic measure of (HRQoL) HRQoL which assesses health using 36 items across eight dimensions. Each dimension is given a score on a 0–100 scale but these scores cannot be used to generate QALYs because they are not based on individual preferences and are not comparable across dimensions.

### The SF-6D

The SF-6D was derived from 11 items of the SF-36 and converted into a six-dimensional health-state classification system, with four to six levels each dimension, allowing for 18,000 unique health states [1]. Dimensions of the SF-6D include: physical functioning, role limitations, social functioning, pain, mental health, and vitality. The SF-6D has a set of preference weights obtained from standard gamble (SG) valuations of a sample of 249 SF-6D health states using a representative sample of the UK population [1]. The SF-6D can be regarded as a continuous outcome measure anchored on a full health-death 1–0 scale where the preference weights range from 0.30 to 1.00 [1]. Since the recent introduction of this preference-based measure, several articles have been published using the SF-6D, e.g. [2–10], and it is anticipated that the application of this measure will continue to grow.

## Methods

### Study Design

This study has two major components: to estimate a system weight for the SF-6D for Portugal, and to compare the results to the UK system weights. For the first component, we replicate the



1098-3015/10/624 624–630

methods followed by Brazier and colleagues [1]. First, we undertook a preference based valuation survey and, second, we modeled the data to estimate a Portuguese system weight for the SF-6D (PT SF-6D). For the second component, we compared the final Portuguese system weight (PT SF-6D) with the UK system weight (UK SF-6D).

An optimal stratified sampling design was used to achieve a representative sample of the general adult Portuguese population that reflects the variability of the Portuguese population in terms of age and sex. The estimated sample size was 140 individuals, with a confidence degree of 95% and a relative precision of 4.60%. Individuals were randomly selected from a survey frame of the Portuguese population over 15 years old including name, address, sex, age and residence. The interviews were conducted in the respondents own homes, between April and October 2006.

Each respondent ranked a set of health states defined by the Portuguese SF-6D and valued these using SG. The Portuguese version of the SF-6D was obtained from the Portuguese version of SF-36 v2 by extracting the 11 items referred in Brazier et al. [1]. To familiarize respondents with the idea of describing health in terms of the SF-6D and as a warm-up exercise, each respondent was first asked to describe their health using the Portuguese versions of the SF-36 v2, the EQ-5D descriptive system, the EQ-5D Visual Analogue Scale and the SF-6D (results not presented). The interviews began with an exercise to rank six health states defined by the SF-6D, plus the best (111111) and the worst (645655) health states defined by the SF-6D and immediate death. Following other researchers [11], we assumed that the ranking exercise is equivalent to the respondent making a series of individual selections from smaller and smaller sets of states.

Respondents were then asked to value the six health states using SG. The SG valuation was based on a variant of the SG using props developed by a team at McMaster [12]. A chance board was used to present the probabilities of both alternatives, both numerically and in a pie chart. Due to its ease of use, the ping-pong (i.e., alternating back-and-forth between high and low values) strategy was employed in the interviews. All interviewers used the interview schedules, response booklets and chance boards suggested by the McMaster team [12]. The research team recruited 10 students to perform the interviews and trained them. No incentive was given to the respondents before or after the interview took place. Following the method used by Brazier et al. [1], after the six SG questions, the respondents were asked a seventh SG question, the form of which depended upon the form they had ranked the "pits" state: better than or worse than immediate death. If they had ranked the "pits" health state as better than immediate death, they were asked to consider a choice between the uncertain prospect of full health (111111) or immediate death and the certain prospect of being in the "pits" state (645655). On the contrary, if they had ranked the "pits" health state as worse than immediate death, they were asked to consider a choice between the uncertain prospect of full health or being in the "pits" state and the certain prospect of immediate death. The response to this SG question was then used to "chain" the health-state values to place them on the full health-death scale (1–0) [1]. The adjustment of the six intermediate SF-6D health-state valuations was made in the following way [1]: $SG_{adj} = SG + (1 - SG)P$, where $SG_{adj}$ is the adjusted SF-6D health-state valuation; $SG$ is the SF-6D health-state valuation; and $P$ is the value of the "pits" state. Respondents were finally asked some socio-demographic questions.

## Selection of the Health States

The SF-6D generates 18,000 different health states. Because it is virtually impossible to value all states, a selection of a number of health states was needed. To compare the results of this research with those of Brazier and colleagues [1], we used the same set of health states as a starting point. Because the minimum sample of health states identified using an orthogonal design is 49 health states, we used the same 49 health states determined by Brazier et al. [1] by applying the orthoplan procedure of the Statistical Package for the Social Sciences (SPSS). Five additional health states were randomly generated to maximize the number of health states valued. Hence, the respondents valued a total of 54 health states, plus the worst health state defined by the SF-6D. These states were classified as mild, moderate, or severe and a stratified sampling method was used to ensure that each respondent was asked to value a set of states from all levels. The severity of the states was assessed by summing the dimension levels, following a recently published approach [13]. The states were then ranked using the score and divided into quartiles to identify the three severity groups. Two health states were then randomly selected without replacement from each group to create nine sets of health states containing six states each. Each health state and set was valued 10 times on average (minimum 9, maximum of 11), adding up to 630 valuations. Table 1 presents the states chosen for the valuation task.

## Modeling Health-State Values

Regression analysis is used to estimate the relationship between the SF-6D and the SG values. These models can then be used to predict health-state values for all states defined by the classification system. Models were estimated at both the individual and aggregate levels because there may be a respondent effect caused by variations between and within respondents [1,14]. First, an ordinary least squares (OLS) regression was used, which relies on the assumption that each individual health-state value is an independent observation. Models with main effects, with interaction effects and with the constant forced to unity were estimated. Breush-Pagan tests revealed heteroscedasticity problems in all

**Table 1** SF-6D Health states valued in the survey

| Number of health states: 54 + worst | |
|---|---|
| 211111 | 144341 |
| 121212 | 224612 |
| 232111 | 534113 |
| 113411 | 633122 |
| 321122 | 235224 |
| 111621 | 334251 |
| 122233 | 414522 |
| 133132 | 432621 |
| 511114 | 431443 |
| 611221 | 443215 |
| 213323 | 622513 |
| 312332 | 625141 |
| 332411 | 315515 |
| 341123 | 115653 |
| 124125 | 523551 |
| 135312 | 642612 |
| 212145 | 323644 |
| 412152 | 545422 |
| 421314 | 614434 |
| 522321 | 531635 |
| 122425 | 631355 |
| 131542 | 115323 |
| 221452 | 211424 |
| 241531 | 112554 |
| 425131 | 611434 |
| 512242 | 645621 |
| 132524 | 645655 |
| 142154 | |

models. To deal with this, all models were estimated using White's heteroscedasticity consistent standard errors.

The model specification at the individual level is:

$$y_{ij} = \alpha + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{r}'_{ij}\boldsymbol{\theta} + \varepsilon_{ij}, \quad (1)$$

where $i = 1,2, \ldots,$ n represents the health states, $j = 1,2, \ldots,$ m represents the respondents, $y_{ij}$ is the adjusted values of the health state $i$ valued by respondent $j$, $\mathbf{x}'_{ij} = (x_{1ij}, x_{2ij}, \ldots, x_{vij})$ is a vector of $v$ dummy explanatory variables referenced to the same unity, in which $x_{vij} = x_{\delta\lambda ij}$ for each level $\lambda$ of dimension $\delta$ of the SF-6D. The $\mathbf{r}'_{ij} = (r_{1ij}, r_{2ij}, \ldots, r_{uij})$ term is a vector of $u$ variables of interaction between the different levels of the attributes, also referenced to the same unity. Further, $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_v)$ and $\boldsymbol{\theta}' = (\theta_1, \theta_2, \ldots, \theta_u)$ are vectors of parameters and $\varepsilon_{ij}$ is a residual variable.

At the aggregate level the model is:

$$y_i = \alpha + \mathbf{x}'_i\boldsymbol{\beta} + \mathbf{r}'_i\boldsymbol{\theta} + \varepsilon_i \quad (2)$$

where $y_i$ is the adjusted aggregated value of the health state $i$.

Second, random effects (RE) models were estimated using generalized least squares (GLS) regressions, allowing for more complex modeling of the variance components observed at both levels of the hierarchy [15]. Models with main effects and with interaction effects are presented. The general RE model is:

$$y_{ij} = \alpha + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{r}'_{ij}\boldsymbol{\theta} + u_j + e_{ij}, \quad (3)$$

where $u_j$ is the respondent-specific variation that is assumed to be random across individual respondents and $e_{ij}$ is an error term for the *ith* health-state valuation of the *jth* individual, assuming that it varies randomly across observations, with $e_{ij} \sim [0, \sigma_e^2]$. Additionally, $Cov(u_j, e_{ij}) = 0$, which means that the health states are randomly allocated to the respondents. The Hausman test was used to test the appropriateness of random effects [16] and the Breush-Pagan Lagrange Multiplier (B-P LM) test was employed to compare the RE GLS and OLS models.

Third, the generalized estimating equations (GEE) approach was used, which is a widely used statistical method in the analysis of longitudinal data in clinical epidemiological studies. It was proposed by Liang and Zeger [17] and Zeger and Liang [18] and is an extension of the generalized linear model algorithm to accommodate correlated data. The GEE approach was used to estimate the RE models with the constant forced to unity, by means of an exchangeable correlation structure. The Hausman test was used to formally test the difference between the fixed effects (FE) and the GEE parameters.

Alternative functional forms were considered to account for the skewed distribution of health-state valuations. A logit, two log-log transformations, and a Tobit transformation were applied. The log-log transformation presented in this article is: $\ln\left[-\ln\left(1 - \frac{1}{2}(SG_{adj} + 1)\right)\right]$, where $SG_{adj}$ is the SG value adjusted to the scale 1–0 (full health-death) as also used in the other models. All functional forms were modeled with RE and GEE. Once again, the Hausman test and the B-P LM test were applied to test the appropriateness of RE.

Because there are potentially thousands of possible interaction effects that could be included in the models, we followed others researchers' approach [1,14] and reduced the interactions by considering a dummy variable (WORSE) for all situations where one or more dimensions are at the worst possible level.

Also following the same approach [1,14], all models were analyzed regarding the expected negative sign and statistical significance ($P < 0.10$) of the coefficients, number of inconsistencies and of its statistical significance, overall fit, and the ability for predicting the SG values for all health states. In common with the literature, we expect coefficients to be negative and increasing in absolute size because the dummy variables represent progressively worse problems on each dimension in comparison to the baseline for that dimension. When a coefficient decreases in absolute size, an inconsistency occurs.

Predictive performance of the models is assessed using the following battery of measures: adjusted $R^2$, mean absolute error (MAE), proportion of health-state values correctly predicted within an error lower than 0.1 ($\%AE < |0.10|$) and than 0.05 ($\%AE < |0.05|$). Goodness of fit assessment includes Akaike's information criterion (AIC) and Bayes information criterion (BIC). Because the GEE method is based on the quasi-likelihood theory, some of the model selection statistics developed under the maximum likelihood theory cannot be applied directly [19]. Therefore, we use an extension of AIC proposed by Pan [20] that uses the quasi-likelihood under the independence model information criterion (QIC) as model selection method. Ramsey's Reset tests (RESET) and tests to the nullity of residuals mean [$t(mean = 0)$] were also applied. Normality of residuals is tested using the asymptotic Jarque-Bera test (JB) and skewness and kurtosis tests for normality. The overall significance of the models is tested using an F test or a Wald-type test ($W(\chi^2)$), as appropriate. All analyses are performed using Stata, release 9.0.

## Results

### Sample

Table 2 shows the main characteristics of the study sample. The sample includes a slight majority of women (53.6%). Age ranges

**Table 2** Study sample and Portuguese population

| | N | % | Portuguese Population over 15 years old (%) |
|---|---|---|---|
| Sample (n) | 140 | 100.0 | 100.0 |
| Female | 75 | 53.6 | 52.3 |
| Married/living together | 86 | 61.4 | 63.5 |
| High Educational Level | 54 | 38.6 | 13.7 |
| Skilled white collar workers | 56 | 40.6 | 15.5 |
| Living in urban areas | 122 | 87.1 | n.a. |
| Income: 1,000€–1,999€ | 54 | 39.4 | n.a. |
| No chronic disease | 90 | 64.3 | 63.4 |
| Mean age (standard deviation) | 42.7 (16.9) | | 46.5 (16.8) |

All values for Portuguese population are based on Portuguese Census 2001 except for the number of individuals with no chronic disease, which is based on 2005/2006 Portuguese National Health Survey.
n.a., not available.

**Table 3** Descriptive statistics for the SF-6D health-states valuations

| State | Va | Mean | SD | Me | Min | Max | State | Va | Mean | SD | Me | Min | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 111621 | 10 | 0.639 | 0.203 | 0.639 | 0.325 | 0.958 | 332411 | 10 | 0.784 | 0.134 | 0.813 | 0.533 | 0.938 |
| 112554 | 10 | 0.737 | 0.218 | 0.816 | 0.278 | 0.978 | 334251 | 11 | 0.576 | 0.270 | 0.618 | 0.058 | 0.948 |
| 113411 | 10 | 0.785 | 0.099 | 0.820 | 0.618 | 0.878 | 341123 | 9 | 0.725 | 0.157 | 0.698 | 0.483 | 0.938 |
| 115323 | 10 | 0.778 | 0.222 | 0.868 | 0.278 | 0.978 | 412152 | 10 | 0.758 | 0.156 | 0.730 | 0.483 | 0.993 |
| 115653 | 10 | 0.608 | 0.161 | 0.665 | 0.325 | 0.798 | 414522 | 10 | 0.364 | 0.244 | 0.336 | –0.013 | 0.675 |
| 121212 | 10 | 0.771 | 0.152 | 0.780 | 0.483 | 0.993 | 421314 | 9 | 0.788 | 0.131 | 0.808 | 0.588 | 0.950 |
| 122233 | 10 | 0.528 | 0.198 | 0.426 | 0.325 | 0.798 | 425131 | 10 | 0.575 | 0.166 | 0.571 | 0.325 | 0.808 |
| 122425 | 10 | 0.746 | 0.223 | 0.835 | 0.278 | 0.988 | 431443 | 10 | 0.589 | 0.209 | 0.618 | 0.278 | 0.808 |
| 124125 | 10 | 0.625 | 0.182 | 0.624 | 0.363 | 0.863 | 432621 | 10 | 0.698 | 0.265 | 0.793 | 0.023 | 0.938 |
| 131542 | 11 | 0.542 | 0.266 | 0.448 | 0.123 | 0.963 | 443215 | 10 | 0.798 | 0.120 | 0.808 | 0.663 | 0.948 |
| 132524 | 10 | 0.725 | 0.136 | 0.683 | 0.528 | 0.903 | 511114 | 10 | 0.666 | 0.235 | 0.618 | 0.348 | 0.958 |
| 133132 | 10 | 0.539 | 0.193 | 0.494 | 0.325 | 0.798 | 512242 | 10 | 0.711 | 0.261 | 0.818 | 0.138 | 0.938 |
| 135312 | 10 | 0.668 | 0.201 | 0.681 | 0.325 | 0.958 | 522321 | 10 | 0.629 | 0.182 | 0.705 | 0.313 | 0.825 |
| 142154 | 11 | 0.560 | 0.252 | 0.448 | 0.258 | 0.963 | 523551 | 9 | 0.733 | 0.132 | 0.698 | 0.588 | 0.913 |
| 144341 | 10 | 0.762 | 0.107 | 0.763 | 0.618 | 0.903 | 531635 | 10 | 0.544 | 0.253 | 0.521 | 0.188 | 0.983 |
| 211111 | 10 | 0.715 | 0.124 | 0.731 | 0.533 | 0.863 | 534113 | 10 | 0.487 | 0.168 | 0.426 | 0.275 | 0.730 |
| 211424 | 10 | 0.812 | 0.208 | 0.865 | 0.278 | 0.988 | 545422 | 10 | 0.596 | 0.290 | 0.683 | 0.123 | 0.878 |
| 212145 | 10 | 0.693 | 0.257 | 0.728 | 0.253 | 0.993 | 611221 | 10 | 0.726 | 0.133 | 0.763 | 0.533 | 0.918 |
| 213323 | 10 | 0.734 | 0.196 | 0.780 | 0.368 | 0.993 | 611434 | 10 | 0.774 | 0.227 | 0.853 | 0.278 | 0.978 |
| 221452 | 10 | 0.660 | 0.262 | 0.785 | 0.138 | 0.913 | 614434 | 10 | 0.217 | 0.314 | 0.048 | –0.088 | 0.640 |
| 224612 | 9 | 0.750 | 0.142 | 0.753 | 0.533 | 0.938 | 622513 | 10 | 0.593 | 0.306 | 0.615 | 0.023 | 0.978 |
| 232111 | 9 | 0.831 | 0.120 | 0.828 | 0.663 | 0.968 | 625141 | 10 | 0.500 | 0.293 | 0.581 | –0.063 | 0.843 |
| 235224 | 10 | 0.620 | 0.169 | 0.609 | 0.368 | 0.863 | 631355 | 10 | 0.658 | 0.289 | 0.689 | 0.023 | 0.978 |
| 241531 | 10 | 0.496 | 0.231 | 0.426 | 0.130 | 0.798 | 633122 | 10 | 0.451 | 0.247 | 0.533 | 0.020 | 0.843 |
| 312332 | 10 | 0.746 | 0.126 | 0.763 | 0.528 | 0.878 | 642612 | 11 | 0.408 | 0.322 | 0.363 | –0.148 | 0.873 |
| 315515 | 10 | 0.710 | 0.179 | 0.725 | 0.363 | 0.968 | 645621 | 10 | 0.674 | 0.269 | 0.713 | 0.278 | 0.978 |
| 321122 | 11 | 0.631 | 0.279 | 0.618 | 0.123 | 0.958 | 645655 | * | 0.278 | 0.363 | 0.300 | –0.500 | 0.950 |
| 323644 | 10 | 0.677 | 0.278 | 0.781 | 0.080 | 0.913 | | | | | | | |

*All individuals valued the "pits" state.
Va Number of valuations; SD Standard Deviation; Me Median; Min Minimum: Max Maximum.

from 15 to 87 years with a mean age of 43 years old (SD = 16.9). The majority of the sample are married or living together as a couple although 30.0% (42) are single. Almost 39% of the sample has a high educational level while 34% (47) with a middle educational level. Respondents are most frequently skilled nonmanual workers (40.6%), living in urban areas (87.1%) and reporting no chronic disease (64.3%). Although 39.4% of the sample had an income between 1000€ and 1999€, 22.6% (31) earned more than 3000€. As was said in the methods section, this random sample is representative of the general adult Portuguese population in terms of age and sex.

## *Health-State Values*

Descriptive statistics for the health states are reported in Table 3. The number of valuations per health state is also presented. The mean health-state valuation is 0.595 (SD = 0.290) and mean health-state values range from 0.217 (614434) to 0.831 (232111). The range of individual values is –0.500 (645655) and 0.993 (121212; 412152; 213323; 212145). Negative values are assigned to some severe health states, but are relatively rare (3.8%).

At the individual level, the distribution of the data is skewed. The skewness coefficient shows the existence of a high negative asymmetric distribution ($g = 1.09$).

## *Modeling the SF-6D*

Several models were estimated but only the most appropriate models are presented here (Tables 4 and 5).

For the mean models (M1 and M2), the majority of coefficients have the expected negative sign. In total, 11 of the 25 coefficients of M1 are significant and there are seven inconsistencies. The inclusion of the interaction term (M2) slightly improved the model because it reduced the number of inconsistencies and increased the number of significant coefficients (12). Compared to a similar model for UK SF-6D in Brazier et al. [1] M1 has fewer significant coefficients (11 vs. 24), more inconsistencies (7 vs. 5) and more positive significant coefficients (2 vs. 0) but better predictive ability according to the proportion of predictions with an absolute error smaller than 5% (65.45% vs. 52.61%). The performance of M2 compared favorably in terms of inconsistencies (6 vs. 5) with the similar UK SF-6D model, which is the model which defines the UK SF-6D system weights [1]. However, our model has fewer significant coefficients (12 vs. 24) and a larger proportion of predictions with an absolute error smaller than 5% (63.64% vs. 51.81%).

The OLS model at the individual level with interaction effects (M4) has 18 coefficients with the expected sign and 18 significant coefficients, but seven inconsistencies. M3 and M4 perform worse than M1 and M2 in terms of predictive ability, having smaller MAE and more observations correctly predicted. However, the comparison of the four models using AIC and BIC indicates that M3 and M4 present better results. The Reset test shows evidence of specification problems for all models. However, the results show that all models are overall significant. The predictions of M3 and M4 are unbiased and prediction errors of M1 and M2 are normally distributed (at the 5% level). Comparisons between models M3 and M4 with the models estimated in the UK study are not possible because their results are not presented in Brazier et al. [1].

The RE models estimated using GLS produce better results than the OLS regression. M5 and M6 have 23 coefficients with the expected negative sign and only five inconsistencies. Furthermore, adding interaction effects to RE (M6) increased the number of significant coefficients from 17 to 20. However, despite the predictions being unbiased and the high number of significant coefficients, there is evidence of specification problems and of non-normality of the residuals. For both models, the majority of the overall variance of the error term can be attributed to the individual effect ($\rho_{M4} = 0.724$; $\rho_{M5} = 0.725$), indicat-

**Table 4** OLS models

| | Aggregated level (mean) | | Individual level | |
|---|---|---|---|---|
| | M1 | M2 | M3 | M4 |
| c | 1.000‡ | 1.000‡ | 1.000‡ | 1.000‡ |
| PF2 | −0.019 | −0.026 | 0.003 | 0.000 |
| PF3 | 0.011 | 0.009 | 0.034* | 0.034* |
| PF4 | −0.022 | −0.027 | −0.021 | −0.020 |
| PF5 | −0.047 | −0.046 | −0.061* | −0.058* |
| PF6 | −0.174† | −0.189† | −0.190‡ | −0.190‡ |
| RL2 | −0.022 | −0.031 | −0.016 | −0.021 |
| RL3 | −0.079* | −0.080* | −0.064‡ | −0.065‡ |
| RL4 | −0.070 | −0.085 | −0.101‡ | −0.109‡ |
| SF2 | 0.024 | 0.018 | 0.003 | 0.005 |
| SF3 | 0.004 | −0.001 | 0.008 | 0.010 |
| SF4 | −0.168† | −0.177† | −0.164‡ | −0.165‡ |
| SF5 | −0.069* | −0.088* | −0.090‡ | −0.096‡ |
| PN2 | 0.079* | 0.077* | 0.070† | 0.068† |
| PN3 | 0.126† | 0.125† | 0.133‡ | 0.140‡ |
| PN4 | 0.053 | 0.053 | 0.052† | 0.055† |
| PN5 | −0.017 | −0.023 | −0.031 | −0.027 |
| PN6 | −0.008 | −0.021 | −0.034 | −0.033 |
| MH2 | −0.030 | −0.041 | −0.042† | −0.047† |
| MH3 | −0.161‡ | −0.166‡ | −0.185‡ | −0.189‡ |
| MH4 | −0.084* | −0.090* | −0.054† | −0.062† |
| MH5 | −0.065 | −0.085* | −0.095‡ | −0.104‡ |
| VT2 | −0.091* | −0.087* | −0.078‡ | −0.076‡ |
| VT3 | −0.082* | −0.075* | −0.074‡ | −0.074‡ |
| VT4 | −0.046 | −0.046 | −0.005 | −0.010 |
| VT5 | −0.014 | −0.029 | −0.059‡ | −0.067‡ |
| WORSE | — | 0.039 | — | 0.025 |
| n | 55 | 55 | 630 | 630 |
| Inconsistencies | 7 | 6 | 7 | 7 |
| MAE | 0.061 | 0.059 | 0.197 | 0.196 |
| $\%AE < \|0.05\|$ | 65.45 | 63.64 | 15.87 | 15.87 |
| $\%AE < \|0.10\|$ | 78.18 | 78.18 | 29.84 | 29.68 |
| $t(mean = 0)$ | (1) | (1) | −0.597 | −0.575 |
| RESET | 5.36† | 4.84† | 5.24† | 2.77* |
| F | 287.14‡ | 273.81‡ | 1369.53‡ | 1322.29‡ |
| JB | 2.954 | 2.926 | 36.39‡ | 36.17‡ |
| AIC | −146.573 | −145.960 | −1223.112 | −1223.776 |
| BIC | −94.383 | −91.762 | −1107.524 | −1103.741 |

(1) Mean error is zero by definition. *$P < 0.10$. †$P < 0.01$. ‡$P < 0.001$.
PF, Physical functioning; RL, Role limitations; SF, Social functioning; PN, Pain; MH, Mental health; VT, Vitality; M1, main effects; M2, interaction effects; M3, main effects; M4, interaction effects.

ing a large degree of unobserved individual heterogeneity [21,22]. Hausman's test suggests that RE, rather than FE, is the appropriate specification. In addition, B-P LM test implies an RE rather than an OLS specification. Comparing M5 with its UK SF-6D counterpart reported in Brazier et al. [1], it is possible to conclude that it presents the same number of significant coefficients and similar explanatory power (adjusted $R^2_{M5} = 0.194$ vs. adjusted $R^2_{\text{Brazier}} = 0.200$). Nevertheless its MAE, is larger than Brazier's model ($MAE_{\text{Brazier}} = 0.073$) and presents a poorer prediction ability ($\%AE < |0.05|_{\text{Brazier}} = 96.50\%$; $\%AE < |0.10|_{\text{Brazier}} = 98.49\%$). Direct comparisons between M6 and a similar UK SF-6D model estimated by Brazier et al. [1] are not possible because the authors include another dummy variable that accounts for all situations where any dimension is at the least severe level.

For M7, all coefficients have the expected negative sign and the vast majority is significant, and the model has only five inconsistencies. The introduction of the interaction term in model M8 slightly improves the results, increasing the number of significant coefficients from 17 to 20 and the variable WORSE is significant and positive. According to QIC, M7 is a better model than M8. Although both models have specification problems, we accept that the regressions are significant overall. The residuals are unbiased, but are not normally distributed. A comparison between M7 and a similar UK SF-6D model [1] shows that, in both, all coefficients are negative, though the number of inconsistencies in M7 is slightly higher (5 against 4) and the number of significant coefficients is lower (17 vs. 26). It also has a poorer prediction ability ($\%AE < |0.05|_{\text{Brazier}} = 96.53\%$; $\%AE < |0.10|_{\text{Brazier}} = 98.32\%$) and a bigger MAE ($MAE_{\text{Brazier}} = 0.078$). Similarly, M8 performs as good as the UK counterpart model [1] in terms of its significant coefficients and performs better in terms of the number of inconsistencies (one less). However, it has poorer prediction ability ($\%AE < |0.05|_{\text{Brazier}} = 96.66\%$; $\%AE < |0.10|_{\text{Brazier}} = 98.55\%$) and a bigger MAE ($MAE_{\text{Brazier}} = 0.076$).

M9 has similar results as M7, where the majority of coefficients have the expected negative sign, less inconsistencies and more significant coefficients, but it has a larger QIC. Once more, it is not possible to compare this model with a similar UK SF-6D model estimated in Brazier et al. [1] because the authors do not report their results.

Table 6 shows the parsimonious consistent models constructed by aggregating levels of each dimension whenever inconsistencies occurred, following other researchers' approaches [23]. For these models, inconsistent estimates in all dimensions have been aggregated to attain consistent scales. The parsimonious consistent model for M7 (M10) is the preferred specification and is paired with the UK SF-6D system weight.

**Table 5** RE and GEE models (individual level: n = 630)

| | RE | | GEE | | Log-Log transformation (GEE) |
|---|---|---|---|---|---|
| | M5 | M6 | M7 | M8 | M9 |
| c | 0.827‡ | 0.817‡ | 1.000‡ | 1.000‡ | 1.000‡ |
| PF2 | −0.041* | −0.040* | −0.050† | −0.050* | −0.065* |
| PF3 | −0.025 | −0.024 | −0.032 | −0.031 | −0.041 |
| PF4 | −0.040* | −0.036* | −0.049* | −0.045* | −0.066* |
| PF5 | −0.045* | −0.042* | −0.055* | −0.052* | −0.069* |
| PF6 | −0.177‡ | −0.179‡ | −0.214‡ | −0.220‡ | −0.320‡ |
| RL2 | −0.026 | −0.028* | −0.031 | −0.034* | −0.065* |
| RL3 | −0.003 | −0.003 | −0.003 | −0.003 | −0.014 |
| RL4 | −0.046* | −0.056† | −0.054* | −0.067† | −0.112† |
| SF2 | −0.031* | −0.033* | −0.038* | −0.042* | −0.055* |
| SF3 | −0.012 | −0.015 | −0.014 | −0.019 | −0.032 |
| SF4 | −0.034 | −0.052* | −0.039 | −0.060* | −0.095* |
| SF5 | −0.057† | −0.066† | −0.069† | −0.080† | −0.114† |
| PN2 | 0.006 | 0.007 | 0.006 | 0.008 | 0.011 |
| PN3 | 0.001 | 0.007 | −0.001 | 0.006 | 0.127 |
| PN4 | −0.049* | −0.047* | −0.061* | −0.060* | −0.074* |
| PN5 | −0.044* | −0.050* | −0.054* | −0.060* | −0.096† |
| PN6 | −0.073‡ | −0.073‡ | −0.090‡ | −0.090‡ | −0.127‡ |
| MH2 | −0.048† | −0.054† | −0.059† | −0.066† | −0.095† |
| MH3 | −0.011 | −0.025 | −0.009 | −0.026 | −0.045 |
| MH4 | −0.057† | −0.059† | −0.070† | −0.073† | −0.101† |
| MH5 | −0.085‡ | −0.092‡ | −0.103‡ | −0.112‡ | −0.180‡ |
| VT2 | −0.043‡ | −0.040* | −0.051† | −0.048† | −0.070* |
| VT3 | −0.031 | −0.026 | −0.036 | −0.031 | −0.059* |
| VT4 | −0.037† | −0.039† | −0.046† | −0.048† | −0.070† |
| VT5 | −0.080‡ | −0.082‡ | −0.097‡ | −0.101‡ | −0.136‡ |
| WORSE | — | 0.033* | — | 0.038* | 0.054* |
| Adjusted $R^2$ | 0.194 | 0.208 | — | — | — |
| Inconsistencies | 5 | 5 | 5 | 5 | 3 |
| MAE | 0.209 | 0.207 | 0.209 | 0.207 | 0.225 |
| $\%AE < \|0.05\|$ | 14.44 | 14.13 | 14.29 | 14.29 | 13.02 |
| $\%AE < \|0.10\|$ | 28.25 | 28.57 | 28.41 | 28.25 | 27.94 |
| $t(mean = 0)$ | 0.000 | 0.000 | 0.000 | 0.000 | 5.419‡ |
| RESET | 14.39‡ | 7.44† | 14.65‡ | 14.46‡ | 15.47‡ |
| $W(\chi^2)$ | 703.05‡ | 711.29‡ | 762.57‡ | 769.61‡ | 786.38‡ |
| JB | 25.72‡ | 25.37‡ | 25.63‡ | 25.28‡ | 0.19 |
| QIC | — | — | 68.069 | 68.231 | 153.387 |

*$P < 0.10$. †$P < 0.01$. ‡$P < 0.001$.
PF, Physical functioning; RL, Role limitations; SF, Social functioning; PN, Pain; MH, Mental health; VT, Vitality.
M5, main effects; M6, interaction effects; M7, main effects; M8, interaction effects; M9, interaction effects.

**Table 6** Parsimonious consistent models (individual level: n = 630)

| | M10-Main effects | M11-Interaction effects |
|---|---|---|
| c | 1.000 | 1.000 |
| PF23 | −0.029 | −0.028 |
| PF4 | −0.047 | −0.042 |
| PF5 | −0.050 | −0.046 |
| PF6 | −0.207 | −0.213 |
| RL23 | −0.012 | −0.013 |
| RL4 | −0.061 | −0.073 |
| SF23 | −0.025 | −0.028 |
| SF4 | −0.051 | −0.071 |
| SF5 | −0.075 | −0.086 |
| PN23 | 0.000 | 0.000 |
| PN45 | −0.049 | −0.052 |
| PN6 | −0.087 | −0.089 |
| MH23 | −0.038 | −0.047 |
| MH4 | −0.066 | −0.070 |
| MH5 | −0.100 | −0.111 |
| VT23 | −0.040 | −0.036 |
| VT4 | −0.041 | −0.043 |
| VT5 | −0.092 | −0.096 |
| WORSE | — | 0.041 |
| MAE | 0.207 | 0.205 |
| %AE < \|0.05\| | 13.81 | 14.60 |
| %AE < \|0.10\| | 29.37 | 28.89 |
| *t*(*mean* = 0) | 0.006 | −0.007 |
| RESET | 19.2* | 15.3* |
| W($\chi^2$) | 741.23* | 751.16* |
| JB | 26.42* | 26.07* |
| QIC | 62.865 | 0.000 |

All the coefficients are significant at the 0.10 level. *$P < 0.001$.
PF-Physical functioning; RL-Role limitations; SF-Social functioning; PN–Pain; MH-Mental health; VT–Vitality.
M10-parsimonious consistent model for M7; M11-parsimonious consistent model for M8.

Figure 1 compares the predicted health-state values of M10 with observed values. It shows that although the tendency to under predict mild health states has been reduced, there is still a problem of under prediction for severe health states. It is clear from the figure that M10 adequately predicts the observed health-state values.

## Discussion

The SF-6D is one of the most widely used HRQoL measurement instruments. Since it was developed in 2002, several authors have studied its properties, its usefulness, as well as its limitations. Many comparisons of SF-6D with other HRQoL measurement instruments have also been published. There have also been some articles reporting the estimation of system weights for the SF-6D for other countries [24,25].

This study reports the results of the survey conducted to estimate the Portuguese system weights for the SF-6D. The results demonstrate that it is possible to estimate preference weights for HRQOL measurement for use in Portugal. The best econometric models adequately predict the health-state values for the general population. Actually, the performance of the RE models with the constant restricted to unity performed better than the FE models. Hence, the recommended PT SF-6D system weights are based on a GEE model with main effects with no inconsistencies (M10). The restriction of the intercept to unity is sustained by the need to generate models for use in cost-utility analyses. Parsimonious consistent models aggregating the levels where there are inconsistent estimates are chosen due to concerns regarding some inconsistent estimates for two levels of the six dimensions. It is worth further research on the under-prediction of the value of the most severe health states observed in some cases. The results were consistent with the valuation studies conducted in UK [1,11]. Socio-demographic data were not included in the models following other valuation surveys reported in the literature [1,14,23,26–28].

Some authors [29] recommend multiplicative models to estimate preference functions for multiattribute health-state classification systems. When compared to additive models, they have the ability to capture interactions between the attributes to a limited degree. The functional form used in the PT SF-6D is a linear additive one, due to the difficulties of estimating and interpreting different functional forms given the nature of the independent variables. The data are explained in terms of a model with one additional term to account for the much greater disutility associated with extreme problems, similar to the model

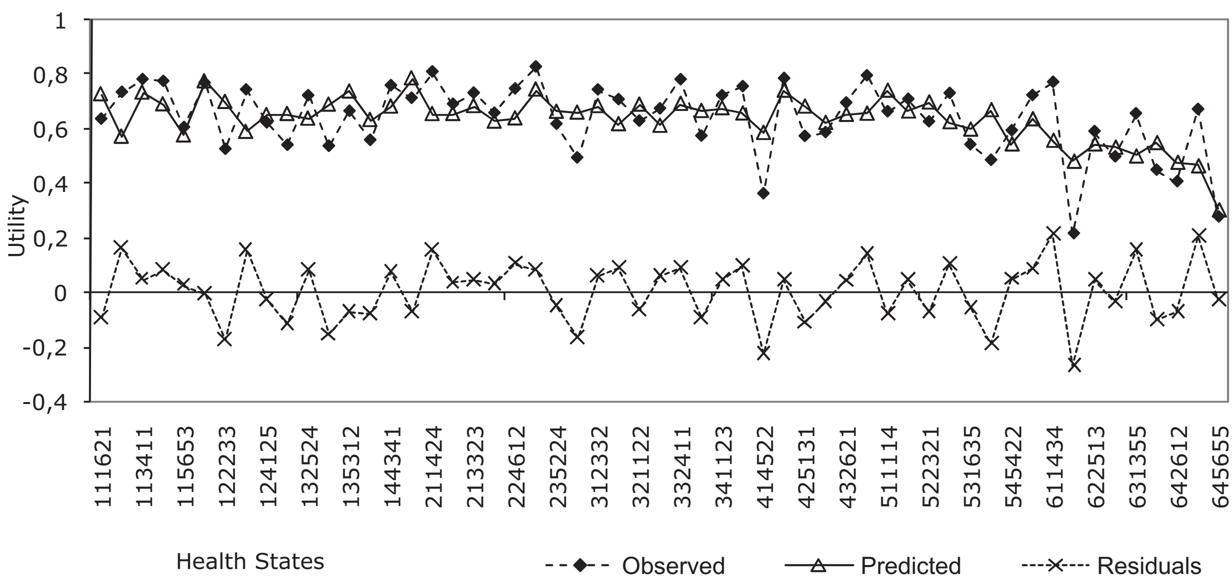


**Figure 1** Actual and predicted health-state valuations for M10.

used for the valuation of the EQ-5D [26]. However, due to the positive sign of the additional term, the preferred specification is a main effects model.

There may be a concern with the small sample size used in the study. The valuation data set is small in terms of both the number of health-state values and the number of respondents, when considering other valuation studies such as the EQ-5D valuation survey [26] or the SF-6D valuation survey undertaken by Brazier et al. [1]. In fact, the results presented here are from 140 respondents, valuing 0.31% of the possible health states generated by the SF-6D. The EQ-5D model was estimated from a sample of 2,997 respondents that valued 43 of the possible health states (17.3%) and the SF-6D was estimated from a sample of 611 individuals that valued 1.4% of the possible health states [14]. A larger sample and an increase in the number of health states valued would probably contribute to a better predictive model. However, it is worth noticing that a random sample was used, therefore allowing for inference over the Portuguese population.

In the future, we intend to use confirmatory analysis and structural equation models to determine the measurement properties of the latent factors underlying the HRQoL of the Portuguese general population and to estimate the magnitude and direction of the interdependent effects among those factors. We also intend to model health-state preference data using rank data and to compare both the rank and the SG approaches.

## Conclusion

The valuation and estimation of a multiattribute preference-based index is a complex task. In this study, we estimated the Portuguese system weight for the SF-6D that can be used in future studies for determining the HRQoL of the Portuguese population.

The authors thank two anonymous referees for their constructive comments and suggestions that have considerably improved the paper.

Sources of financial support: Lara Ferreira and Luis Pereira were beneficiaries of fellowships (SFRH/BD/25697/2005 and SFRH/BD/36764/2007, respectively) of the Foundation for Science and Technology, Portugal. No other funding was received.

## References

1 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002; 21:271–92.
2 Marra C, Woolcott J, Kopec J, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and The HAQ) in rheumatoid arthritis. Soc Sci Med 2005;60:1571–82.
3 Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13:873–84.
4 Ferreira P, Ferreira L, Pereira L. How consistent are health utility values? Qual Life Res 2008;17:1031–42.
5 Lamers L, Bouwmans C, van Straten A, et al. Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ 2006;15:1229–36.
6 Marra C, Esdaile J, Guh D, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med Care 2004;42:1125–31.
7 Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. Health Econ 2005;14:1169–89.
8 Pickard A, Simon J, Jeffrey A, Feeny D. Responsiveness of generic health-related quality of life in stroke. Qual Life Res 2005; 14:207–19.
9 Stavem K, Frøland S, Hellum K. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Qual Life Res 2005;14:971–80.
10 McTaggart-Cowan H, Marra C, Yang Y, et al. The validity of generic and condition-specific preference-based instruments: the ability to discriminate asthma control status. Qual Life Res 2008;17:453–62.
11 McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. J Health Econ 2006;25:418–31.
12 Furlong W, Feeny D, Torrance G, et al. Guide to Design and Development Health-State Utility Instrumentation. Working Paper Series. Ontario: McMaster University Press, 1990.
13 Brazier J, Roberts J, Platts M, Zoellner Y. Estimating a preference-based index for a menopause specific health quality of life questionnaire. Health Qual Life Outcomes 2005;3:13.
14 McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the HUI2 descriptive system: results from a UK survey. Health Econ 2005;14:231–44.
15 Rabe-Hesketh S, Everitt B. A Handbook of Statistical Analyses Using Stata (4th ed.). Boca Raton, FL: Chapman & Hall/CRC, 2007.
16 Baum C. An Introduction to Modern Econometrics Using Stata (1st ed.). College Station, TX: Stata Press, 2006.
17 Liang K-Y, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.
18 Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
19 Cui J, Qian G. Selection of Working Correlation Structure and Best Model in GEE Analyses of Longitudinal Data. Comm Statist Simulation Comput 2007;36:987–96.
20 Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57:120–5.
21 Jones A. Applied Econometrics for Health Economists—A Practical Guide (2nd ed.). Oxford: Office of Health Economics, 2007.
22 Jones A, Rice N, Bago d'Uva T, Balia S. Applied Health Economics (1st ed.). London, UK: Routledge Taylor & Francis Group, 2007.
23 Brazier J, Roberts J. The estimation of a Preference-Based Measure of Health from the SF-12. Med Care 2004;42:851–9.
24 Lam C, Brazier J, McGhee S. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. Value Health 2008;11:295–303.
25 Brazier J, Fukahara S, Roberts J, et al. Estimating a preference-based index from the Japanese SF-36. J Clin Epidemiol. 2009;62:1323–31.
26 Dolan P. Modelling valuations for EuroQol health states. Med Care 1997;35:1095–108.
27 Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 health survey. J Clin Epidemiol 1998;51:1115–28.
28 Dolan P. Modelling valuations for health states: the effect of duration. Health Policy 1996;38:189–203.
29 Stratmann-Schoene D, Kuehn T, Kreienberg R, Leidl R. A preference-based index for the SF-12. Health Econ 2006;15:553–64.